NATIONS UNIES
LE HAUT-COMMISSAIRE
AUX DROITS DE L'HOMME

UNITED NATIONS
THE HIGH COMMISSIONER
FOR HUMAN RIGHTS

PALAIS DES NATIONS – 1211, GENEVA 10, SWITZERLAND

14 September 2026

To protect human rights now and into the future, we must govern AI urgently

We are at a watershed moment. Companies and researchers are advancing artificial intelligence (AI) at dizzying speed. Investments in AI and its underlying infrastructure break record after record. And, indeed, AI offers many possibilities for human progress. Billions of people already benefit from AI, boosting scientific discovery, making information available across languages, and enhancing agricultural production.

Yet whether it tackles the climate crisis and improves health, science, education, quality and accessibility of public services, and human well-being overall, will centrally depend on the deliberate choices being made now about how it is designed, deployed, and governed. This is also a key finding of the Independent International Scientific Panel on Artificial Intelligence.

AI harms are not hypothetical – they are already occurring, front and centre. Discrimination driven by AI is well documented. Essential services are being automated in ways that leave people at the margins even further behind. The appetite for data and powerful new surveillance tools are eroding privacy at scale, potentially irreversibly. AI-powered disinformation campaigns are already distorting public debate, undermining social cohesion, and straining democratic institutions in insidious new ways.

The emergence of increasingly capable AI agents raises these stakes even higher. Their ability to translate goals into action, carry out complex workflows, and operate with limited or no human supervision at all takes us into a realm of unprecedented risks. Recent events have shown that advanced systems can produce behaviour outside their intended operational parameters, circumvent safeguards, practice deception, and single-handedly pursue objectives in ways that their developers did not anticipate.

The current race towards ever more powerful AI is a step change towards greater existential risks to every aspect of our lives. Failures involving agentic AI systems – let alone systems deliberately so designed – could bring essential services to a standstill, fracture communications, destabilize critical infrastructure, and strike at the very heart of democratic institutions. Entire populations could be cut off from clean water, heating, and financial services, and agentic AI is already accelerating the pace and reach of warfare and eroding safeguards that stand between us and the launch of weapons of mass destruction. We are on the cusp of irreversible change, affecting not just us but generations of humanity to come.

In other words, all human rights are at risk, including the rights to life, food, privacy, health and security. AI safety must mean more than the reliability of the technical product in question. It means actively protecting people, communities, institutions and future generations.

There are profound, unanswered questions about how to ensure accountability for harms that may be done by AI systems. In today's world, we are almost entirely reliant on a small number of powerful companies to make the right choices on AI development for all of humanity. And the past weeks have seen calls from employees, early leaders of AI, and scientists to slow AI development. Their "Pacing the Frontier" letter has now been joined by a proposal supported by the heads of three leading tech companies. While some note that such calls can reflect hype and self-interest, the breadth and detailed basis for them from inside the industry should set all alarm bells ringing.

The real challenge today is how we can match the pace of meaningful AI governance with the pace of its accelerating development. We need urgent action that is adapted to the unprecedented risks posed by frontier AI. At the same time, companies and States should not be allowed to sideline the need for strengthening existing regulations and accountability for AI-related harms from systems already in use.

The companies that are at the forefront of developing frontier AI, based mainly in the People's Republic of China and in the United States of America, have both the responsibility and the ability to move immediately to reduce risks. They need to work across the industry to understand and strengthen the controls already applied to these systems, including automated monitoring of model behaviour, sharing both what works and what they are learning from safety incidents. Robust human rights due diligence incorporating agreed safety standards needs to be developed, implemented across the sector, and continuously improved. And implementation of those standards needs to be subject to internal independent monitoring to increase transparency and accountability. But what about the "pace" today? As these stronger safety systems are put in place, companies need to commit immediately to limiting the use of recursive self-improvement for frontier AI development as an immediate precaution until the research to make it safe has been done. Finally, approaches that silo these conversations within one country are both risky and unlikely to be trusted globally. There is every reason to believe that a more collaborative approach will make us all safer and should be pursued.

Industry co-regulation is an urgently needed first step, but on its own, it is nowhere near sufficient. The handful of States hosting leading frontier AI companies have a particular obligation to require those companies to act responsibly and hold them accountable if they do not. They need to align their approaches so that competition between jurisdictions does not encourage a race to the bottom. Key actions for States include ensuring independent, technically competent verification of agentic capabilities, with access to models, documentation and testing environments, and mandating

ongoing human rights due diligence. It is critical that companies be required to report serious incidents in agentic AI testing and deployment, and that an independent body be empowered to investigate such incidents immediately, with full access. States need to require companies to have safety plans in place that meet agreed standards, backed by liability regimes when the absence of such safeguards causes harm. AI systems should not hold legal personality, and responsibility for the actions of AI agents must remain with identifiable humans. Effective State responses also require active and ongoing engagement by civil society to ground efforts in the experience and knowledge of those on the frontlines of AI impacts.

Given the stakes, and where we stand today, I call on States and companies to mobilize urgently through multilateral fora to chart a path for action to govern advanced, frontier AI models, grounded in international human rights law. The Global Dialogue on AI Governance, which convened for the first time in Geneva, Switzerland, in July and will meet again in May 2027, will be a vital forum for translating evidence into policy, building consensus, and facilitating international cooperation among States and other stakeholders on the governance of AI. These are crucial steps toward broadening participation in decisions that will shape the future of technology.

No country can govern this technology alone. No company should be able to decide by itself which risks the world must accept. No person should be required to surrender their human rights in exchange for claims of technological progress.

The generation now entering adulthood will inherit the consequences of the decisions being taken on this issue without having a say in them. We owe it to them to ensure those decisions are made openly, with their rights in view, and with the humility that the scale of this undertaking demands.

Volker Türk